



Digital Scholarship at the Library of Congress

User demand, current practices, and options for expanded services

March 17, 2020

Authors

Chairs

- Grant Harris, Chief, European Division
- Abigail Potter, Senior Innovation Specialist, Digital Strategy Directorate
- Kate Zwaard, Director, Digital Strategy Directorate

Working Group

- David Brunton, Chief, Platforms Division
- Susan Garfinkel, Digital Reference Specialist, Researcher & Reference Services Division
- John Hessler, Cartographic Specialist, Geography & Map Division
- Christa Maher, Digital Projects Coordinator, Digital Collections Management & Services Division
- Jaime Mears, Senior Innovation Specialist, Digital Strategy Directorate
- Nicole Saylor, Head of the Archive, American Folklife Center
- Stephanie Stillo, Subject Collections Specialist, Rare Book & Special Collections Division
- Candice Townsend, Supervisory Librarian, Researcher & Reference Services Division

Contents

I. Introduction	2
II. Digital Scholarship at the Library of Congress.....	3
Affordance: A Guiding Principle.....	4
Case Studies: Current Practices and Unmet Demand.....	5
1. Geographic Information System for CRS.....	5
2. Text Analysis of U.S. Elections.....	5
3. Machine Learning and Classified Ads.....	6
4. Mapping Spanish Music.....	6
5. Mobilizing Haitian Journals.....	7
6. Bulk Downloads from Subscription Databases.....	8
7. Ask a Librarian	8
8. A Web Archives Pilot.....	9
III. Recommendation: Prioritize Digital Collection Readiness.....	10
Computational Readiness of Digital Collections.....	10
Online and Accessible for Computing.....	10
Online but Optimized for Browsing.....	13
Accessible Only Onsite.....	14
Section Recommendations	14
IV. Recommendation: Build Institutional Capacity	16
National Digital Initiatives.....	16
The Library of Congress Geospatial Hosting Environment.....	17
Community of Practice	18
Section Recommendations	21
V. Recommendation: Expand User Services.....	22
Section Recommendations	22
VI. Conclusion	23

I. Introduction

The Library of Congress created the Digital Scholarship Working Group in March 2017 to identify how the Library enables digital scholarship, documents current user demand, and recommends models to expand support. Members of the group represented eight divisions from within three service units in the Library, bringing together a variety of experiences and perspectives for this paper.

The Working Group delivered its report to Library management in December of 2018. In the interest of sharing the analysis performed and recommendations, the Library made this public version by removing internal information and editing for clarity. Please note that, except for these changes and updates to the introduction, some footnotes, and numbers, **the report has not been revised** to reflect changes in our environment, developments in the field, or the evolution in our own thinking.¹ This report predates the Library's latest Strategic Plan² and Digital Strategy.³ This report offers recommendations to expand support for digital scholarship at the Library of Congress, some of which have been implemented, but does not speak to funding sources or the tradeoffs that might be necessary for execution.

The Working Group analyzed a sample of recent digital scholarship requests at the Library (see the case studies starting on page 5). The lessons from these case studies reveal how the Library can expand digital scholarship support to meet current needs. The Working Group's recommendations fall broadly into three categories:

- **Invest in collection readiness**—The Library could make many more digital collections usable for computation if they were machine readable and accessible as bulk downloads. The Working Group recommends continuing the work of migrating existing online collections to common web platforms, clarifying rights and privacy restrictions, and investing in research on how best to provide computational access to the Library's heterogeneous digital archival collections. Further, the Working Group recommends supporting existing machine-readable interfaces and providing more bulk-data downloads. Providing data transformations and documentation will make collections more usable for digital scholarship projects.
- **Build institutional capacity for digital scholarship**—The Working Group recommends helping Library staff develop an understanding of the types of digital scholarship requests we can currently accommodate with our collections and service infrastructures, and how investment would allow us to meet more of the existing needs. Further, the Working Group recommends developing skills, ethical frameworks and practices, and knowledge across the Library that can augment support for digital scholarship reference inquiries. This capacity building could be further supported by creating a community of practice around digital scholarship at the Library of Congress.
- **Expand user services**—As the evidence base for the kinds and complexity of user requests and options for service models grow, the Working Group recommends the Library consider additional investments to enhance support for users using digital methods with the Library's collections.

¹ The working group members' organization titles on the cover sheet were current at the time of internal publication.

² "Enriching the Library Experience," Strategic Plan, Library of Congress, <https://www.loc.gov/strategic-plan/>.

³ "Digital Strategy for the Library of Congress, FY 2019-2023," Library of Congress, <https://www.loc.gov/digital-strategy/>.

II. Digital Scholarship at the Library of Congress

Current practices, user demand, and options for expansion

“Digital Scholarship’ is defined as any scholarly activity that makes extensive use of one or more of the new possibilities for teaching and research opened up by the unique affordances of digital media.”

-Digital Scholarship, University of Washington⁴

For several decades, an international and diverse group of specialists has combined the potential of digital collections with innovative tools to develop new ways of accessing, analyzing, and sharing information. Early digital scholarship included textual, geospatial, network, and computational analysis as well as new models for curation and the creation of digital editions.⁵ These tools and methods have led to everything from the spatial “imaginary” of ancient Athens and the 1893 World’s Columbian Exposition, to striking visualizations of early modern European literary networks and Depression-era Farm Security Administration photographs. They have also re-contextualized digital collections onto public websites, where public historians, other scholars, and students have combined library digital collections with other resources to tell new stories. Digital scholarship now features recognized methodologies and tools, and practitioners in nearly every discipline.

Equally exciting are the emerging audiences for digital scholarship. As digital content has continued to democratize information by reducing traditional barriers to research, the ready availability of digital tools, combined with new entrées to rich historical and cultural content, inspires creativity in artists, journalists, historians, genealogists, citizen scientists, educators, students, and life-long learners.

The growth of digital scholarship, digital humanities,⁶ and other public-facing digital projects has led collecting institutions to consider their roles and responsibilities. The term “collections as data” has emerged in the past to describe the goals and work of libraries, archives, and museums seeking to provide content and services to support digital work. In 2016, the Institute of Museum and Library Services funded the *Collections as Data: Always Already Computational* project, which produced *The Santa Barbara Statement on Collections as Data*. The *Santa Barbara Statement* noted that “By conceiving of, packaging, and making collections available as data, cultural heritage institutions work to expand the set of possible opportunities for engaging with collections.” In an effort to call attention to the need for libraries to contend with some of the risks inherent in computational methods such as machine learning, the authors emphasized that “collections as data stewards are guided by

⁴ “About Digital Scholarship,” University Libraries, University of Washington, <http://www.lib.washington.edu/digitalscholarship/about>.

⁵ For additional context, see Chris Alen Sula and Heather Hill, “The Early History of Digital Humanities,” DH2017, ADHO, <https://dh2017.adho.org/abstracts/347/347.pdf>.

⁶ “Broadly construed, digital humanities is the use of digital media and technology to advance the full range of thought and practice in the humanities, from the creation of scholarly resources, to research on those resources, to the communication of results to colleagues and students.” Dan Cohen, “Defining Digital Humanities, Briefly,” *Humanities* (blog), March 9, 2011, <https://dancohen.org/2011/03/09/defining-digital-humanities-briefly/>.

ongoing ethical commitments.”⁷ The Library of Congress also hosted events around the collections as data topic, focusing on stewardship and use models in 2016,⁸ and impact in 2017.⁹

Throughout 2017, the Digital Scholarship Working Group documented requests to use Library collections as data, examined the readiness of collections for computational use, surveyed the Library’s digital scholarship efforts to date, and researched how peer institutions are preparing their workforce to meet this challenge. This effort has revealed just how much momentum digital scholarship is gaining at the Library. A greater investment in services, infrastructure, and training is needed, not only to meet current demand but also to position the Library to remain a vital resource in future decades. With the continuing development of computing resources that enable digital scholarship—including hardware, software, digital content, research-specific tools, and both qualitative and quantitative computational methods—as well as the concurrent development in practices of collaboration and publishing, two common needs have emerged. The first is access: the easy availability of digital and digitized content in high-quality, broadly usable formats. The second, affordance, is the opportunity to conduct digital research using the tools and methods best suited for a chosen context or research community.

Affordance: A Guiding Principle

As described by Donald A. Norman in his 1988 book, *The Design of Everyday Things*,¹⁰ an affordance is like a doorknob: a point of contact enabling interaction, where the design of the thing encourages a particular use—doorknobs invite turning, for example. If we consider the library as a single entity designed for use, we can identify many of its features, from infrastructure to services to staffing, as affordances that encourage, invite, and support particular uses. Affordances of a library might include well-lit reading rooms that enable onsite reading and Interlibrary Loan programs that support use of materials offsite. The affordances of a library can extend from lists of titles in alphabetical order to the classification systems that invite browsing of materials according to topic. Creating effective affordances that support scholarship requires attention to the full research ecosystem, as well as the cultural contexts in which the materials have meaning and the range of methods that the library is designed to support. For digital scholarship this includes not just hardware, software, digital content, and computational methods, but also the needs, knowledge, actions, and activities of both digital researchers and library staff who, through their explorations, collaboratively turn that digital content into understanding.

Libraries have long worked to provide not just access to, but affordances for, particular uses of their content. Cataloging, catalogs, and OPACs; bibliographies, finding aids, and pathfinders; in-person and online research guidance; reading rooms for research and display cases for exhibits; and physical walls that enable legal use of rights-restricted content: these are just some examples of ways that libraries support particular uses of collections.

⁷ “The Santa Barbara Statement on Collections as Data,” Already Always Computational—Collections as Data, GitHub, <https://collectionsasdata.github.io/statement/>.

⁸ September 2016 talks from “Collections as Data: Stewardship and Use,” Videos, Library of Congress, <https://labs.loc.gov/meta/videos/>. The livestream from “Collections as Data: Impact,” July 25, 2017, video, 7:17:24, <https://www.youtube.com/watch?v=OJWMHzgCu3c>. Since the delivery of this report, the Library has hosted additional Collections as Data-themed events: <https://labs.loc.gov/events/>.

⁹ In 2018 the Andrew W. Mellon Foundation funded a second round of the *Collections as Data* project that is distributing \$30,000 - \$80,000 to 12 teams to develop models that support implementation and use of collections as data. “Collections as Data—Part to Whole Call for Proposals,” GitHub, <https://collectionsasdata.github.io/part2whole/cfp/>.

¹⁰ Donald Norman, *The Design of Everyday Things*, New York: Basic Books, 1988. See especially pp. 87-91.

Most recently, “data” as a type of digital library content, are receiving more attention. Data are pieces of information in aggregate, often gathered into some sort of readable entity—a table, chart, book, database—but more processing is nearly always required to make data accessible for a particular use. Considering the ways that data are offered, and the affordances provided for that data, is an important part of library work. Geo-location, visualization, text mining, topic modelling, and image analysis are just some of the data-based methods that are gaining broad popularity, as data sets and open-source tools to investigate them develop.

Case Studies: Current Practices and Demand

The case studies below provide examples of digital scholarship reference requests the Library received through consultations in reading rooms, informal consultations with researchers, and other channels. These ran from general advice about tools and techniques to specific requests for Library collections data in alternate formats (e.g., multi-page downloads, textual transcriptions). The questions came from a broad range of users, including graduate students and professors, journalists, librarians, artists, authors, engineers, and lawyers. Similar questions make their way to Library reference staff, curators, technology specialists, or developers through contacts with researchers and colleagues at the Library and peer institutions. Increasingly, visiting scholars at the John W. Kluge Center arrive with specific research goals that incorporate digital approaches or requirements. The launch of LC Labs (labs.loc.gov) in September 2017 further raised the Library’s profile of digital scholarship offerings, making both new and existing resources more discoverable. It also generated heightened interest in digitally driven, and sometimes experimental, uses of Library content.

1. Geographic Information System for CRS

At a 2017 Reference Roundtable, Congressional Research Service (CRS) librarians shared that they are getting an increasing number of requests from Congressional users for Geographic Information System (GIS) visualizations—maps presenting data needed for decisions on urban planning, infrastructure design, or policy analysis. They also noted that the visualization products they create are shared among peers, generating more requests. Another staff member confirmed that CRS is currently exploring strategies to help meet this need, including the launch of a geospatial GIS portal in partnership with the Geography & Map Division (G&M). The portal functions as a data pipeline to securely deliver CRS and G&M GIS products directly to members of Congress and their staff.

Lessons learned: Demand for CRS-produced map visualizations is high and expected to grow. The Geospatial Hosting Environment (GHE) is a high-value addition to CRS’s services to Congress. It can be assumed that the GHE will be popular with other Library users; a specific example of researcher use of the GHE is given below.

2. Text Analysis of U.S. Elections

The Library’s web archiving program preserves websites and serves them as single browsable pages on loc.gov. In the past few years, the Library has received an increasing number of requests for bulk delivery of web archives, a feature not currently available through the existing interface. In one example, a graduate student contacted the Web Archiving team to request bulk download of the United States Elections Web Archive for all available candidate sites and election cycles. The researcher wanted to use text analysis on the corpus to analyze whether candidates’ positions on certain issues were different before or after a specific historical event.

The Library stores web archives in the Web ARChive (WARC) file format. These files are organized by capture, not by collection (e.g., a given WARC file will have websites from many collection captured on a certain day), with onsite-only or embargoed websites included with the ones that are open access. Rights issues aside, because

the WARCs are not limited to the target information, the size of the files transferred would not be useful to most users.

Lessons learned: There is researcher demand for off-site bulk access to web archives. Investments in a workflow to create access derivatives (i.e., files with a subset of web archives) and file delivery services would enable the Library to serve these requests where there are no access restrictions.

3. Machine Learning and Classified Ads

The John W. Kluge Center¹¹ (hereafter: Kluge Center) at the Library of Congress hosts a diverse community of scholars at various stages of their careers from all over the world to use the Library's collections for their research and to share their expertise with the Library and the broader Capitol Hill community.

The research topic for a scholar in residence as a Kluge Digital Studies Fellow¹² was how people have bought and sold items using informal markets. She worked with the Library's *Chronicling America* collection of historic newspapers to identify instances of fraudulent ads and bartering, comparing these to contemporary ads on Craigslist. "The first step would be developing an archive of classified ads, since the digitized newspapers don't allow you to search on that specific section," she wrote to a Serial & Government Publications Division librarian. *Chronicling America* offers full-text searching of millions of historic newspaper pages, but there is no capability to select just classified ads as a subset. While conducting research at the Library, the assistant professor took a Software Carpentry class at the Library of Congress to learn more about the Python programming language to help her analyze the data once it was compiled.

Lessons learned: The digitized and freely available content could not be readily limited to the subset required by the researcher, despite consultations and the researcher's willingness to learn programming skills. This case study reflects the significant investment necessary to create item and article-level categorization, which could be enabled by exploring automated and semi-automated processes.

4. Mapping Spanish Music

Another Kluge Fellow wanted to geographically visualize traditional Spanish music in the American Folklife Center's Lomax Family Collection using the Library's Geospatial Hosting Environment (GHE). She taught herself how to use the necessary software and created an interactive map. A policy question emerged: would the Library publish the research products of people who are officially affiliated with the Library but are not staff members on its GHE platform? The Fellow's term expired before that question was settled, and her project was shared on labs.loc.gov.¹³

Lessons learned: The Library's GHE includes an approachable, geospatial story-telling tool that is usable for those of all levels of programming skills. However, a clear policy is needed on access to software accounts, as well as criteria and policies for hosting and publishing on labs.loc.gov or anywhere else.¹⁴ This case study

¹¹ "About the John W. Kluge Center," Library of Congress, <https://www.loc.gov/programs/john-w-kluge-center/about-this-program/>.

¹² More about the Kluge Center's Digital Studies Fellowship, with periodic open application periods: <https://www.loc.gov/programs/john-w-kluge-center/chairs-fellowships/fellowships/kluge-fellowships-in-digital-studies/>.

¹³ Ascensión Mazuela-Anguita, "Alan Lomax's Journey Across Spain (1952-53)," ArcGIS StoryMaps, <http://www.arcgis.com/apps/MapTour/index.html?appid=5769e4432d8b4fae841b1a170a4aed98>.

¹⁴ The Library established a policy for intern, fellow, and volunteer Story Maps accounts in 2018, after the internal release of this report.

reflects the readiness of some of our digital collections, but simultaneously the lack of policy clarity around non-staff use of Library tools and hosting on loc.gov and labs.loc.gov.



Figure 1. Screenshot of a Story Maps visualization showing Lomax's fieldwork in Spain.

5. Mobilizing Haitian Journals

A university professor came to the Library for its rare, complete run of the Haitian journal, *Revue de la Société Haïtienne d'Histoire, de Géographie et de Géologie*, wishing to compile and publish an annotated bibliography optimized for access at slower internet speeds and available in both the U.S. and Haiti.

As an alternative to her practice of using a spreadsheet to compile metadata and citations, which she found cumbersome, she was interested in moving to an open-source static site generator to share this work, especially if it would support a feature to solicit comments and feedback. The researcher consulted with National Digital Initiatives (NDI)¹⁵ staff and a visiting software development librarian from George Washington University¹⁶ on the technical aspects of this project. The software development librarian helped the professor transition from the spreadsheet to an open-source citation-management program designed to export and format data, support collaboration, and integrate with other platforms. To make that possible, the software development librarian wrote code to reformat the spreadsheet data for ingestion into the citation software.

After reviewing publishing options with the researcher, the software development librarian recommended using a CMS plug-in with the citation software as a publishing platform. Although she had limited web publishing experience, the researcher was willing to learn what was necessary to complete her project. Running a combination of two established and hosted applications offered ease of publication and required minimal

¹⁵ National Digital Initiatives (NDI) is the Library of Congress division that launched LC Labs and is now part of the Digital Strategy Directorate.

¹⁶ NDI hosted Laura Wrubel on a three-month sabbatical from GWU Libraries to explore how to support digital scholarship.

technical skills. Once the data were loaded, the researcher connected with staff at her home university library who worked with her to update and launch her project, first listed as an experiment at <http://labs.loc.gov>¹⁷ and now available at <http://rshhgglab.com/>.

Lessons learned: The assistance from the visiting software librarian on this project was instrumental to its success. She offered technical and domain expertise to facilitate a digital-scholarship project using Library collections that is a unique contribution to an area of study. This case study reflects collections in a ready state and a request achievable with existing infrastructure. However, the staff and services provided were only available while the software librarian spent her sabbatical at the Library, and it is worth considering the scalability of this level of support. A point of promise is that this scholar had home institution support when her time at the Library ended, which could present a potential connection point. The project was ultimately successful, but without additional staff, partnerships, or handoffs to other libraries, not repeatable.

6. Bulk Downloads from Subscription Databases

The Library received a request to download a subset of in-copyright items for which the Library provides on-site access through subscriptions to commercial databases. For the purpose of this example, think of a bulk download of, for a particular time period, the sports sections of modern national newspapers, the table of contents of French fashion magazines, or the last journal article in a list of American medical journals. This material had not been digitized by the Library, and the subscription databases did not offer bulk download through the user interface. The sought-after material was under copyright restriction, and licensing agreements might further limit the Library's flexibility in how it could serve the material. Unfortunately, though there is a desire to work with our vendors to get bulk or raw access to content to which the Library subscribes, the request for this content could not be fulfilled during the necessary timeframe.

Lessons learned: This is an example of a digital scholarship request that is challenging to fulfill due to legal and contractual obligations and technical cost. In some cases, although the researcher may assert that their educational use of in-copyright material is legal, the Library has other legal or contractual obligations that limit our ability to serve it in the manner requested. This case study reflects the need to continue to explore licenses for subscription databases that allow bulk downloading, and the advantage of pre-planning digital scholarship projects whenever possible.

7. Ask a Librarian

Patrons send digital scholarship research requests to the Library's "Ask a Librarian" online reference service, a selection of which the Working Group extracted from a correspondence system using full-text search. The search did not reveal a comprehensive list of these requests: relevant terminology is quite varied and not all researchers recognize that computational methods are needed to answer their questions. Questions were found from as far back as 2013, including requests to run text and sentiment analysis on a text archive; download the entire MARC data set (which later became a service provided by the Library¹⁸); access a sortable index of a U.S. president's correspondence in order to make an interactive timeline using XML; and to provide assistance in the use of text analysis tools with a particular collection that is available as HTML. The Library's ability to assist individual patrons who contact us via "Ask a Librarian" or by email depends on a variety of factors that can include the expertise of librarians who encounter and route the questions; the availability of technical staff to

¹⁷ See the labs.loc.gov description and link to the RSHHGG Lab, <https://labs.loc.gov/experiments/rshhgglab/>.

¹⁸ "MARC Distribution Services (data set)," Library of Congress, <https://www.loc.gov/cds/products/marcDist.php>.

respond to patrons when this work would need advanced IT skills; and the feasibility of fulfilling, in a reasonable turnaround time, one-off requests for specific data in a specific format.

Lessons learned: Some patrons who pose digital scholarship queries via “Ask a Librarian” do have their needs met, but others end up rethinking their research plans or looking elsewhere for guidance or content that the Library is not readily able to provide. Staff capacity could be expanded in two ways. There is an opportunity to help public-facing service staff, such as reference librarians, expand their ability to answer digital scholarship inquiries and to connect them with other librarians with digital scholarship skills. Additionally, technical staff who occasionally work with patrons directly could benefit from training in reference work and quick evaluation of the level of effort necessary to answer the user’s need.

8. A Web Archives Pilot

As part of the 2016 *Library of Congress Lab: Library of Congress Digital Scholars Lab Pilot Project Report*, Michelle Gallinger and Daniel Chudnov engineered a successful technical pilot outlining how the Library might provide collections as data at-scale. For their project, Gallinger and Chudnov sought collections that were digital, had clear access restrictions, and allowed use in some capacity. A quote from their report:

“[With] guidance from Library of Congress staff, we chose to work with a subset of the Web Archives collection. This selection ... allowed us to consider a file format (WARC) requiring specialized processing and which has met with growing scholarly interest in analysis using contemporary tools for distributed processing of large-scale datasets.... Browsing through the Library of Congress web archive collections via Wayback lets us see individual pages of handfuls of websites at distinct points in time, but this one-link-at-a-time paradigm is a mismatch for the capacity of a historian armed with a computing cluster wishing to perform content or network analysis over millions of web pages containing billions of links.... Given these issues and the increasing value of access to web archives collections as data and at scale not otherwise being served currently by the Library of Congress, we believe this selection of web archive materials illustrates the gap in access to digital collections which a Lab might be positioned to fill.

To that end, we successfully transferred over five terabytes of WARC files with web pages collected from links in syndication feeds from major news sites the Huffington Post and the *Detroit Free Press*, comprising material including HTML, images, web scripts and style documents, and linked videos from YouTube which the Library collected in support of its broader collections. Because this particular set of materials were crawled with the specific intent to complement broader crawls, rather than to stand on their own as collections for public access, they are not otherwise scheduled for cataloging or public access and have not been studied on their own in any way. They have been collected on a daily schedule for over a year, allowing us to proxy a scholarly role by looking at patterns in how the materials evolve over this time period.”¹⁹

Lessons learned: This case study used collections that were (almost) ready for use, a workflow that was developed to fulfill this request, and staff who were able to give the appropriate technical guidance and contextual information so that the report’s authors could conduct the pilot. A caveat to this case study is that the authors, both former staff members with technical expertise, were very familiar with the Library’s collections and policies.

¹⁹ Michelle Gallinger and Daniel Chudnov, “Library of Congress Lab: Library of Congress Digital Scholars Pilot Project Report,” (2016), 12-13, http://digitalpreservation.gov/meetings/dcs16/DChudnov-MGallinger_LCLabReport.pdf.

III. Recommendation: Prioritize Digital Collection Readiness

“Stripping digital collections down to core components could render everything old new again in terms of what libraries might offer to the humanities research community.”
- 2016 ALA “Special Report: Digital Humanities in Libraries”²⁰

Computational Readiness of Digital Collections

Since 1990, with the American Memory pilot project, the Library has worked to digitize its collections and, shortly after, to collect born-digital material. By 2018, the Library managed approximately 410 million digital files²¹—15.9 petabytes of content—comprising a tremendous corpus of images, manuscripts, audiovisual materials, notated music, books, archived websites, maps, newspapers, and more. As described in the case studies, some users have expectations that are not matched to current access systems or available derivatives. Offsite patrons contacting the Library via “Ask a Librarian,” for example, sometimes assume that the entire contents of the Library of Congress are available to them in digital form. Additionally, they might have in mind a specific format or derivative that may or may not exist or would be expensive to transfer in bulk, like corrected full-text transcription, high-quality image scans suitable for printing, high-quality audio files, or video files suitable for displaying on large screens. And at the same time that browsable and searchable content has become a commonplace expectation, a newer set of digital affordances are increasingly expected by Library patrons working in areas of digital scholarship. These may include batched sets of metadata, image files, OCR texts, comma-delimited data sets, machine-readable code for use in a growing set of off-the-shelf digital tools for textual analysis, visualization, and more.

The following section gives an overview of the current state of readiness of the Library’s digital collections at two levels of affordance: basic access that makes the materials available on the Internet or in reading rooms, and computational readiness that makes digital materials accessible and easy to use with digital tools and methods.

Online and Accessible for Computing

Bulk Downloads

The Library provides two processed bulk-download derivatives:

- The National Digital Newspaper Program provides metadata, page images, and OCR-derived text from digitized American newspapers.²²
- MARC Distribution Services provides access to nearly 25 million MARC records as a data set.²³

Both of these data sources have been used by the public in diverse creative ways. Digital humanists have been using the Chronicling America data for the past decade to research such topics as the history of newspaper

²⁰ Stewart Varner and Patricia Hswe, “Special Report: Digital Humanities in Libraries,” *American Libraries*.
<https://americanlibrariesmagazine.org/2016/01/04/special-report-digital-humanities-libraries/>.

²¹ From the “Frequently Asked Questions” of the Digital Collections Management website, updated for the public version of this paper in February 2020,
<https://www.loc.gov/programs/digital-collections-management/about-this-program/frequently-asked-questions/>.

²² “Chronicling America OCR Data,” <http://chroniclingamerica.loc.gov/ocr/>.

²³ “MARC Distribution Services (data set),” Cataloging Distribution Services, Library of Congress,
<https://www.loc.gov/cds/products/marcDist.php>.

publishing and the American West,²⁴ how newspapers reported epidemics,²⁵ and even to create automation tools that can be applied to other data sets, such as Elizabeth Lorang's image-based classifier for detecting poetic content in historic newspaper collections.²⁶ The National Endowment for the Humanities has used Chronicling America data as an engagement tool, running challenges²⁷ for members of the public to create web products demonstrating its potential.

In May 2017, the Library offered, for the first time, bulk download of 45 years of bibliographic data.²⁸ Quite quickly, people found interesting applications for the data. Later that month, National Digital Initiatives used them as an engagement and educational tool at an event called Hack-to-Learn.²⁹ Matt Miller,³⁰ a developer at the New York Public Library, created a list of all of the unique titles at the Library of Congress.³¹ The list is roughly 9 million titles long and fully zoomable, so that each title can be read individually, depicting the scale of the Library's holdings in an accessible and engaging format.



Figure 2. Visualization of unique titles in the Library of Congress MARC record release by Matt Miller.

Ben Schmidt, Assistant Professor of History at Northeastern University, plotted all of the publishing and MARC record dates to show a history of how the catalog grew over time. His analysis also uncovered what he

²⁴ Geoff McGhee, "Journalism's Voyage West," The Bill Lane Center for the American West, Stanford University, <https://west.stanford.edu/research/works/journalism-s-voyage-west>.

²⁵ Tom Ewing, Samah Gad, and Naren Ramakrishnan, "Infectious Reporting: Network Analysis of the 1918 Influenza Epidemic Using Historical Newspapers," (2013) http://mith.umd.edu/sharedhorizons/wp-content/uploads/InfectiousReporting_SharedHorizons_VT_April2013.pdf.

²⁶ Elizabeth Lorang, "Developing an Image-Based Classifier for Detecting Poetic Content in Historic Newspaper Collections," D-Lib Magazine, July/August 2015, Volume 21, Number 7/8, <http://www.dlib.org/dlib/july15/lorang/07lorang.html>.

²⁷ "NEH Announces the Winners of the Chronicling America Data Challenge National Endowment for the Humanities," July 27, 2016, <https://www.neh.gov/news/press-release/2016-07-25>.

²⁸ "Library Offers Largest Release of Digital Catalog Records in History," Library of Congress, <https://www.loc.gov/item/prn-17-068/library-offers-largest-release-of-digital-catalog-records-in-history/2017-05-16/>.

²⁹ Mears, Jaime, "Hack-to-Learn at the Library of Congress," *The Signal*, <https://blogs.loc.gov/thesignal/2017/06/hack-to-learn-at-the-library-of-congress/>.

³⁰ Between the internal release of this report and its public release, Matt Miller became a Library of Congress staff member. The work referenced here was completed at NYPL.

³¹ Miller, Matt, "List Vertigo," May, 2017, <https://s3.amazonaws.com/loctitles/index.html>. Accessed 19 Dec 2017.

at first thought were errors, but after publishing his findings and methodology on his blog, catalogers responded, explaining some of the trends in his visualization with great detail and informing some of his assumptions, enabling a compelling cross-disciplinary conversation. Through this collaboration, the Library has new access to data about the history of its contributions in this area.

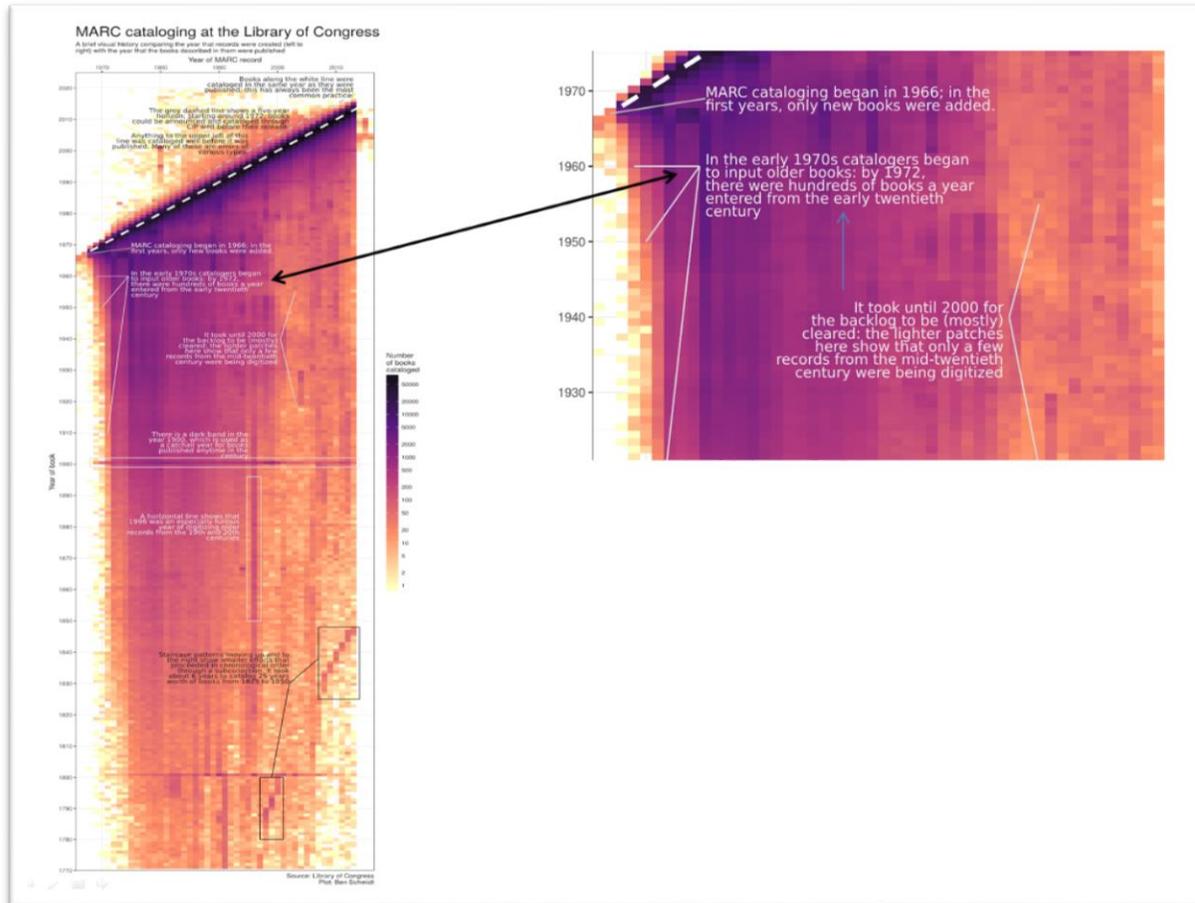


Figure 3. Dates of publication and dates of MARC records, visualized together by Ben Schmidt.³²

Application Programmer Interfaces (APIs)

There are additional collections for which bulk data are available to those with the technical skills to write software that can download information from either an API or a website.

The Library’s newly released digital collections are available in the Project One³³ framework on loc.gov. In addition, over the past five years, the Library engaged in the labor-intensive process of migrating collections from older online presentations, such as American Memory, Global Gateways, Chronicling America, and the National Jukebox. This work will continue for years to come.

³² Ben Schmidt, “A brief visual history of MARC cataloging at the Library of Congress,” <http://sappingattention.blogspot.com/2017/05/a-brief-visual-history-of-marc.html>.

³³ “Project One” is the internal Library of Congress name for the search and discovery system and interface at <https://www.loc.gov/collections>, to which many older web presentations have been migrated.

Digital collections in the Project One interface on loc.gov are accessible via the Library’s JSON API. Users can access it by affixing “?fo=json” to the end of many loc.gov URLs (see <http://www.loc.gov/pictures/?fo=json>). Additionally, XML sitemaps are available for more than 400 digital collections,³⁴ providing information such as item URLs, date of last modification, and frequency of updates.³⁵ Similarly, <http://id.loc.gov/>³⁶ has a number of formats available for crawling a large number of data sets, but bulk download of data is not generally available.

LC Labs has published a list of Library of Congress APIs, bulk downloads, and tutorials called LC for Robots³⁷ to help users explore machine-readable access to Library collections data. This has helped researchers with the requisite technical skills to make use of these tools to build new ways to access the Library’s collections that are available via the API (such as ProPublica data fellow Kate Rabinowitz’s @LoC Map Bot³⁸).

Online but Optimized for Browsing

Digital content “above sea level” is accessible, but optimized only for interactive browsing. Some of the complexities in the way collections are structured, for example, can present a challenge to automated interaction. As of February 5, 2020, the Library has more than 400³⁹ digital collections online and publicly accessible from <https://www.loc.gov/collections/>, with more added every month, and still more available on other presentation platforms. These collections include multiple content formats, from photographs and audio-visual materials to archived websites. Each digital collection on the loc.gov presentation platform includes an overview of and contextualization for the digital content and collection items, which consist of the digital objects and their descriptive metadata.

While the Library has made great progress in standardizing the production of digital collections, there remains variation, largely because of content migrations from other web presentations.

Some digital collections (often those that have migrated from legacy presentation formats, such as American Memory) have added-value articles and essays to give context; others may only have an “About this Collection” summary, a “Rights and Access” statement, and featured items. Some collections have multiple file formats available for download, including text from human transcription or OCR, TIFF master images, derivative jpgs, thumbnails, jpeg2000, and item-level PDFs for users to download a multi-image item in one click. Others may have a single format available for download (such as a PDF or text file), or content that is available for streaming only—often, but not always, due to rights restrictions. Additionally, in the cases where downloads are permitted, there is often no easy way to download in batches; thus, to create a computable data set, researchers have to download items one by one and aggregate them.

The descriptive metadata displayed on loc.gov and used in the online search functionality is derived from multiple sources using the Extract-Transform-Load (ETL) process. ETL ingests and indexes descriptive records and identifies content files to display. There are different ETL processes based on the source and delivered format of descriptive data, including MARC records from the ILS, METS records for some special

³⁴ “Digital Collections Sitemap,” Library of Congress, <https://www.loc.gov/collections/sitemap>. Retrieved Feb. 5, 2020.

³⁵ For example, “Franklin Pierce Papers sitemap,” Library of Congress, <https://www.loc.gov/collections/franklin-pierce-papers/?c=1000&sp=1&fo=sitemap>.

³⁶ “LC Linked Data Service: Authorities and Vocabularies,” Library of Congress, <https://id.loc.gov/>.

³⁷ “LC Labs, LC for Robots,” Library of Congress, <https://labs.loc.gov/>.

³⁸ Kate Rabinowitz, “LoC Map Bot,” <https://twitter.com/LoCMapBot>.

³⁹ Some Prints & Photographs Division collections may not be included in this number.

formats (e.g., music, newspaper, and web archives), “non-MARC” XML,⁴⁰ web crawls, and WordPress. While the outcome of the varied ETL processes provides, at a minimum, a standardized JSON expression of descriptive metadata for each item, access to descriptive record data in other formats largely depends on the source of the data and its native format. For instance, items whose descriptive records come from the Library’s catalog allow us to automate derivatives in MARC XML, MODS, and Dublin Core, while “non-MARC” XML records provide access only to the JSON.

Accessible Only Onsite

The Library digitizes an estimated 5 million files a year, prioritizing rights-free Americana and inclusive of popular international collections and preservation needs. The Library is also working toward increasing capacity, moving from a yearly release average of 5-20 newly digitized collections to 30-35. Additionally, the work of migrating from older presentations to Project One means more digital collections are available via the API every year.

Yet, there is opportunity to make even more digital collections accessible for computational research with technical investment, optimization in understanding the rights associated with content, and additional development of policy pathways.

The requisite rights-related work is substantial and includes interpreting restrictions from the donor or rights holder, licenses, acquisition agreements, and copyright status. Work to streamline the investigative labor involved would help to minimize the amount of staff intervention necessary for access to these materials. Further, there is an opportunity to make clear policy decisions about access levels, such as, under what circumstances Library staff can transfer digital content in bulk to offsite users.

Archival digital collections can present a significant challenge to computational analysis. Sometimes the arrangement of digital collection files makes them virtually unusable without the help of staff analysis and interpretation of arrangement and description. For instance, multiple gifts from a donor can include duplicate files, inconsistent file structures, and mismatches with donor-provided descriptive metadata. Still other digital collections could be made more accessible after it has been determined which methods and presentations are most useful to users, while keeping in mind security and scalability concerns and the complexity in serving heterogeneous collections—a process which is ongoing.

Section Recommendations

A small portion of the Library of Congress’s digital and digitized collection is online, ready for computation, and ready for users. The following steps are recommended in order for digital scholarship researchers to get full value from the digital collections.

Make more collections available

- Streamline the path between accessioning of digital collections and their availability via the API;
- Determine areas of investment that could allow more bulk access to digital archival collections;
- Continue content migrations from legacy presentations to presentations that afford API access;
- Investigate possibilities for enabling onsite computational analysis to digital collections that are not publishable to the open web, where ethical and legal restrictions allow.

⁴⁰ “Non-MARC” is a metadata encoding schema used at the Library for metadata exchange between applications, developed before the creation of MODS.

Develop features for computational use

- Dedicate support and maintenance for API and bulk data access;
- Create a “data format” on loc.gov (i.e., loc.gov/data/);
- Create scaffolding to make data sets more usable by publishing version numbers, cryptographic hashes, and the method of creation or capture;
- Automate transformations from native formats to user-friendly formats where helpful (i.e., repackaging web archives from WARCs to access formats);
- Explore both the risks and the feasibility of processing collection items to extract data, (e.g., additional OCR, color analysis, handwriting recognition, sound recognition, etc.) within an established ethical framework;
- Provide more support for metadata enhancement and correction;
- Consider and document the ethical and legal implications of providing broader computational access to collections.⁴¹

Provide more information

- Determine the level of investment necessary to, on a large scale, clarify usage restrictions and appropriate uses (i.e., rights status, donor agreements, contractual conditions), which optimally:
 - Provide rights information at the item level;
 - Provide rights statements that users find easy to understand;
 - Continue to add new rights assessments for items that don’t have them.
- Provide more documentation aimed at computational users on what is available at loc.gov for which collections, e.g., file formats, OCR, multi-image downloads, IIIF implementation;⁴²
- Document the “Extract-Transform-Load” process that collects metadata from multiple sources and schemas to enable search and discovery on loc.gov.

⁴¹ For more about the ethical and professional standards for creating, sharing, and using collections as data, see the 2019 OCLC research report, “Responsible Operations: Data Science, Machine Learning, and AI in Libraries,” by Thomas Padilla <https://www.oclc.org/content/dam/research/publications/2019/oclcresearch-responsible-operations-data-science-machine-learning-ai.pdf>.

⁴² See, for example, “About American Memory: Technical Information,” Library of Congress, <https://memory.loc.gov/ammem/about/techIn.html>.

IV. Recommendation: Build Institutional Capacity

“My promise to the library community is that the Library of Congress will share its resources. We want to be part of the ecosystem.”

Dr. Carla Hayden at OCLC ARC, 2017

The Library of Congress has been a leader in providing access to, and affordances for, its collections. In 1966, it initiated the Machine-Readable Cataloging program (MARC) that is still in use today. In 1990, the American Memory project began digitizing unique materials to distribute on disks and shifted in 1993 to providing access through the World Wide Web. Soon after this came the National Digital Library Program, followed by the National Digital Information Infrastructure Preservation Program. Other examples include web archiving, the Flickr Commons program, Chronicling America, and the ongoing digitization of millions of individual library items covering all subjects, genres, and original formats. The Library continues to pioneer digital initiatives. The loc.gov development team also continues to add new functionality for digital content on the website, such as using the IIF framework. These activities support some digital scholarship needs, but could be expanded.

National Digital Initiatives

The Library created the National Digital Initiatives (NDI) Division to increase the visibility and impact of its digital collections. NDI’s “Collections as Data” symposium in fall 2016 convened programmers, artists, entrepreneurs, activists, researchers, and librarians to discuss ways of finding new meaning in digitized and born-digital content. Thousands of people participated either in person or through the live stream, reflecting the interest of the greater library community for dialogue on this topic. In May 2017, NDI hosted a hackathon called *Hack-to-Learn*, where research tools were demonstrated using collection data sets to faculty, librarians, students, and interested users. Collections included 52,000 transcribed index cards of Phyllis Diller’s jokes, provided by the Smithsonian Institution, and the Library of Congress’s recently released MARC record data set. In a similar effort to provide training in analyzing and managing digital collections, NDI also hosted a Software Carpentry workshop,⁴³ inviting 40 staff members from the Library of Congress, the DC Public Library, and various federal libraries for hands-on learning in the programming language Python, version-control software Git, and the command-line interface Bash.



Figure 4. The Phyllis Diller team at *Hack-to-Learn*, May 17, 2017.

⁴³ “Software Carpentry Workshop,” GitHub, <https://oulib-swc.github.io/2017-02-15-loc/?loclr=blogsig>.

In December 2016, NDI commissioned a report investigating how libraries and other research centers have developed services that use computational analysis, design, and engagement to enable new kinds of discovery and outreach. In their report, *Library of Congress Lab*,⁴⁴ the authors made recommendations for ways the Library could support digital scholarship with appropriate staff, robust computing resources, and close consultation with librarians, including establishing a lab to serve the scholars in residence at the Library’s Kluge Center. The authors also conducted a pilot program that demonstrated a workflow and created a budget for securely serving web archives in bulk.

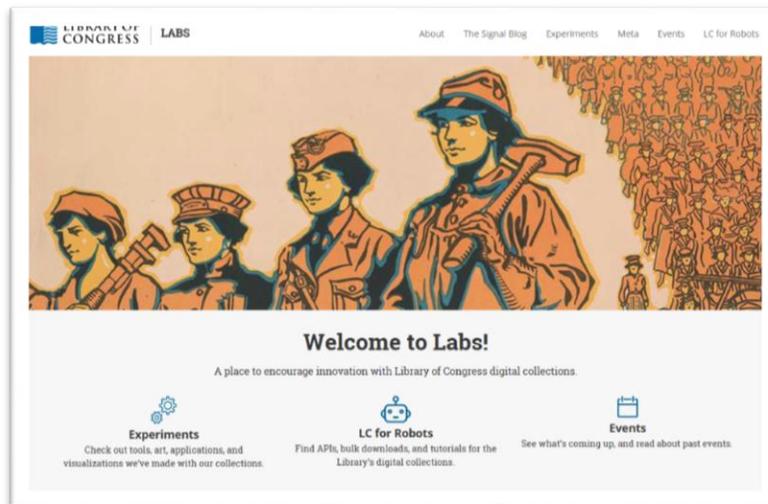


Figure 5. Screenshot of labs.loc.gov.

In the fall of 2017, NDI launched the website labs.loc.gov—an entryway to computational use of the Library’s collections, intended to engage the public and elicit feedback. The site includes apps, educational resources, an “LC for Robots” page that lists resources like guides to using the Library’s APIs and data sources, and a showcase of projects using Library collections. The site also features a crowdsourcing platform that allows users to identify and transcribe image captions from the *Chronicling America* Newspaper Collection.

The Library of Congress Geospatial Hosting Environment

Several years ago, the Geography and Map Division (G&M), in conjunction with the Congressional Research Service (CRS), launched a geospatial data infrastructure—the Geospatial Hosting Environment (GHE), which serves G&M, CRS, the Law Library, and Congressional offices across the Capitol campus.

The GHE consists of three portals and a geospatial data discovery tool that allows geospatial data to be securely transferred, acquired, and stored. Geospatial analysis and custom data visualizations can also be produced and delivered through both desktop and cloud-based Geographic Information Systems software.

To date,⁴⁵ approximately 296 Congressional and Senate offices have made use of the software and the portals to produce geospatial analysis and maps for both legislative policy and constituent outreach. Both CRS and G&M

⁴⁴ Gallinger and Chudnov, 2016.

⁴⁵ As of the original internal publication of this report in 2017.

use the software to provide custom geospatial analysis, data visualizations, and mapping directly to Congress and reading room patrons.

As an outgrowth of the GHE, the Library completed a pilot program using cloud-based story mapping software, which allowed several divisions to produce digital humanities mapping applications for the web. The pilot was quite successful, producing 10 cloud-based applications that feature the Library's collections. The program is now open to all Library staff, with procedures and policies well documented in the staff intranet.

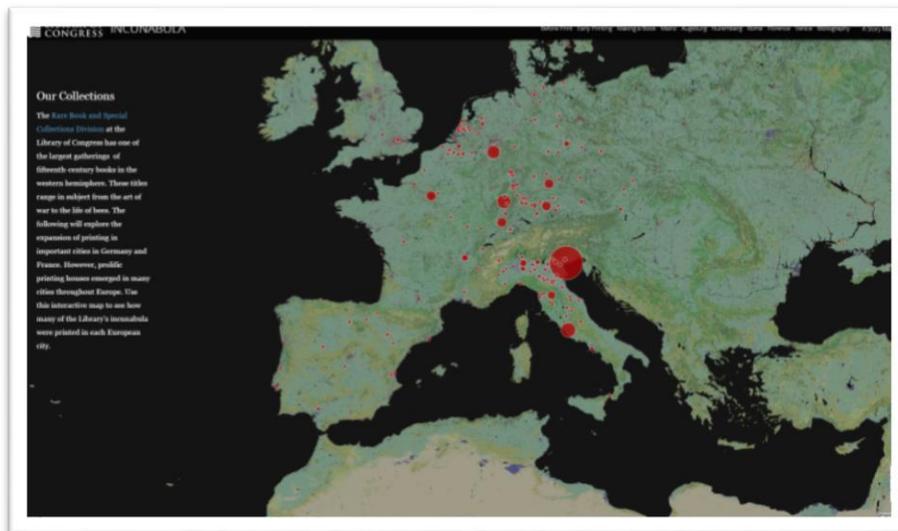


Figure 8. Screenshot of the Geospatial Hosting Environment visualizing the Incunabula collection on a map.

Community of Practice

Although the Library supports several exciting digital initiatives, an opportunity exists for developing a comprehensive strategy to meet the growing demand for digital scholarship. It is certainly not alone in this regard. According to a 2016 ALA special report on digital humanities, advancement in digital scholarship efforts nationally has been uneven but constant across the library landscape, with some type of support now available at

“Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly.”

most libraries. Only 17% of libraries responding to the ALA survey reported no digital scholarship services at their institutions. About 19% of libraries that consider themselves “actively engaged” had built expansive DH centers, and 21% had created special digital humanities librarian positions or were cross-training existing staff.⁴⁶

Maintaining an active and engaged digital scholarship community at the Library can be accomplished by supporting staff in building digital scholarship literacy. As information resources shift, grow, and adapt at an ever-increasing rate,

⁴⁶ Stewart Varner and Patricia Hswe, “Special Report: Digital Humanities in Libraries,” *American Libraries Magazine*, January 4, 2016, <https://cdr.lib.unc.edu/concern/articles/9k41zg75t> (pre-print). More context about this article is available here: Stewart Varner, “Um ... about that American Libraries article we wrote,” January 4, 2016, <https://stewartvarner.com/2016/01/um-about-that-american-libraries-article-we-wrote/>.

Library staff continue to develop expertise across multiple domains of knowledge. These include new developments within subject specializations; methods for digitization, preservation, and the effective creation and use of library metadata; new avenues for interacting with patrons via a growing range of social media outlets; and research techniques for exploring library content using digital tools and methods developed by the scholarly community.

One mechanism for building staff capacity is to establish “communities of practice” within the Library. As explained by learning theorists Etienne and Beverly Wenger-Trayner, “Communities of practice are groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly.”⁴⁷ They note three characteristics that are crucial to a working community of practice: 1) a shared domain of interest; 2) a community that pursues its common interests in that shared domain; and 3) a practice, defined as a repertoire of specific goals, procedures, and resources developed and shared within the community about the domain of interest.

In order to better understand the communities of practice model, the Digital Scholarship Working Group consulted with several national libraries with active digital scholarship programs.

The British Library

The Digital Scholarship Working Group at the Library of Congress conducted a series of interviews with the Digital Scholarship Department (DSD) at the British Library (BL) to better understand how national libraries are meeting the demands of 21st-century scholarship. The DSD team provides support for BL staff and serves as a valuable resource for researchers: the team provides both training and transformative access to BL collections.

The DSD raises awareness about digital scholarship on BL’s campus in several ways, e.g., monthly digital scholarship reading groups, ‘Hack and Yack’ sessions that focus on new areas of research, and monthly one-on-one meetings with lead curators. In addition, the DSD provides an impressive range of courses designed specifically for reference and curatorial staff. These include:

- **Digital Scholarship 101**—Explores how information technology has transformed research. It offers an overview of concepts, tools, and methods that define digital scholarship today.
- **Information Integration: Mash-ups, APIs, and the Semantic Web**—Introduces the fundamentals of information integration and sharing, from web mash-ups to APIs and Semantic-Web linked open-data technology, and how these are being used to connect collections online.
- **Cleaning Up Data**—Teaches BL staff how to normalize data and helps with projects. For example, curators and reference staff enlisted the help of DSD to normalize the date formats in cataloging records.
- **Crowdsourcing and Collaborative Projects**—Provides an overview of crowdsourcing in memory institutions, illustrating best practices and providing guidance on implementation.
- **Data Visualization for Analysis in Scholarly Research**—Provides an overview of key concepts in visualization. It also introduces freely available visualization tools that help scholars understand, analyze, and present large-scale data sets.

The six DSD staff members who work with curatorial departments also seek out possible joint projects across divisions.

⁴⁷ Etienne and Beverly Wenger-Trayner, “Introduction to Communities of Practice: A Brief Overview of the Concept and Its Uses,” 2015, <http://wenger-trayner.com/introduction-to-communities-of-practice/>.

The DSD office also provides direct assistance to researchers, who submit queries through a website. The DSD team also works with BL reference staff to help them understand which questions that come in through the usual reference channels they can help with. The DSD seeks to continually improve its practices. For example, in the formative years of the DSD, the researchers they worked with relied on a model used for doctoral students collaborating with the BL, in that they needed a badge to access internal-only resources. Now, in phase 2 of the DSD's development, researchers simply need a visitor's pass and are taken to a "hot desk" that allows them to query databases from within the firewall. They are also allowed to have storage devices, which will soon be able to accommodate tens of terabytes of content. The DSD also allows some researchers to connect remotely, working with a virtual machine that prevents them from downloading the data to their personal computer.

The DSD employs two dedicated research software engineers, one working on internal collaborations and the other consulting with researchers. In 2016 the DSD accepted 50 researcher consultations, a fraction of the demand for services. Of those 50 consultations, staff successfully resolved 40 of the questions using the existing digital collections without creating special software, which was necessary for the remaining 10 questions. A recent DSD survey shows that the number of potential DH questions posed to the BL per year will likely be above 500 in the near future; the DSD is looking at ways to meet that demand without adding many new staff members.

The DSD reaches out to about 20 British universities regarding digital scholarship. The DSD also has a Collaborative Doctoral Program, which is unique to the United Kingdom.

Other National Libraries and U.S. Academic Libraries

Other examples of national libraries supporting digital scholarship include the Royal Library of the Netherlands (Koninklijke Bibliotheek, or KB), which has an interactive website⁴⁸ and a digital humanities team of five staff members, including two software engineers. The KB has provided more than 10 million searchable historical pages, most of them in Dutch, of newspapers, books, journals, and radio bulletins.

In 2008, Denmark formally merged three institutions, with the full name of "The Royal Library, the National Library of Denmark and the Copenhagen University Library," into the largest library in the Nordic countries. Denmark also created a consortium for digital humanities called DigHumLab,⁴⁹ which includes four universities and all but the 'Royal' part of the three-institution library, to form a "national, distributed research infrastructure that integrates and promotes tools, digital resources, communities, and opportunities to Danish researchers in the humanities and social sciences."

To get a sense of affordances for digital scholarship in academic libraries across the United States, the Association of Research Libraries (ARL) has profiled about twenty DH centers that provide formal support, including at the universities of Illinois, North Carolina, and Virginia. ARL's April 2016 overview noted the following:

"To support digital scholarship a number of ARL institutions founded digital humanities and digital scholarship centers within their libraries as early as the mid-1990s, but since 2000 the number has increased, with more than 25 such centers founded since 2005—at least 8 opened their doors in 2014 alone. A survey of ARL

⁴⁸ "Digital Humanities," Koninklijke Bibliotheek, <https://www.kb.nl/en/organisation/research-expertise/digital-humanities>.

⁴⁹ DIGHUMLAB, <https://dighumlab.org/>.

members, conversations with library deans and directors, and the interest in the CNI/ARL Planning a Digital Scholarship Center Workshop make it clear that many of those institutions that do not yet have a center plan to create one within the next one to five years. Those that do not plan to create such a center already offer decentralized or distributed support for digital scholarship across their campuses that include work and staff within their libraries.”⁵⁰

The Library can model the strategies of the BL and KB by consciously nurturing an expanded community of practice built around the specific needs of digital scholarship at the Library of Congress that could include not only nodes of collaboration around specific projects but also an ongoing and dedicated program of training (at various levels of formality).

Build Institutional Capacity: Section Recommendations

The Working Group recommends the Library consider funding a concerted effort to create a community of practice around digital scholarship at the Library of Congress:

- Offer introductory and higher-level training and professional development opportunities for staff, including temporary details in different departments;
- Building institutional capacity must include the continued development of ethical and professional values and standards for using and sharing collections;
- Develop performance targets and other mechanisms that include support for digital scholarship work;
- Cultivate partnerships with other organizations for knowledge, skills, tools, and staff exchanges, especially focus on building relationships with digital humanities centers. Seek out or lead collaborative approaches to support digital scholars and projects;
- Host workshops and events about digital scholarship. Offer support and services to digital scholarship researchers and other users;
- Continuously highlight the developments and opportunities of digital scholarship at the Library.

⁵⁰ Rikk Mulligan, “Digital Scholarship Support in ARL Member Libraries: An Overview,” Association of Research Libraries, April 19, 2016, <http://www.arl.org/focus-areas/scholarly-communication/digital-scholarship/digital-scholarship-support/3978>.

V. Recommendation: Expand User Services

To make the most use of the Library's investment in its digital assets, the Digital Scholarship Working Group recommends that the Library of Congress consider different service models to further increase the level of support for researchers and staff in computational inquiries. Service models can range widely in scale of investment and are not mutually exclusive. For example, the Library can train existing staff to provide reference support for computational uses, as is currently available for Chronicling America. The Library could also fund additional staff dedicated to this purpose. At the largest scale, the Library could seek funding to create a space that encourages collaborative investigation and testing of tools and workflows to support these ends.

Based on peer examples such as the British Library Labs,⁵¹ recommendations from Gallinger and Chudnov's report, and the case studies mentioned in the previous section, the Working Group imagines a grand vision of cross-disciplinary services that support patrons, Kluge scholars, subject matter experts from partner universities, curatorial staff, and reference staff interested in exploring and interpreting the Library's digital collections and metadata. An investment in a physical center and full-time employees could also improve our ability to provide reference training, data programming, tool development, and outreach. Services would include staff answering technical and reference questions, assisting in research projects, and instructing through formal programming and informal consultations. Staff would also be responsible for developing workflows, techniques, and tools to make accessible options for data delivery and for use of the Library's collections.

Additional infrastructure in an established space could support a robust virtual environment needed by staff and users to work with collections at scale, especially as the Library develops its on-site-only digital collections. These "hot desks," such as those in the Digital Scholarship Department at the British Library, could allow patrons to analyze terabytes of content, while respecting security and rights concerns.

Section Recommendations

The following recommendations, based on peer examples such as the British Library Labs and the KB Lab, would enhance the Library's ability to support computational research with its collections. It is important to note that these recommendations (like many of the recommendations in this report) are not currently funded and represent a menu of ideas for consideration when the Library evaluates priorities. Whether envisioned as a Center or as a number of iterative projects, expanded user services could include:

- Digital scholarship tools and services, including reference support for onsite and remote users;
- Programming and regular training to expand user services;
- Space for in-person collaboration and digital exhibits.

⁵¹ British Library Labs, <https://labs.bl.uk/>. Other examples include: Digital Scholarship Lab, University of Richmond (<http://dsl.richmond.edu/>); Digital Scholarship Lab, Brown University (<https://library.brown.edu/dsl/>); CESTA (Center for Spatial and Textual Analysis), Stanford University (<https://cesta.stanford.edu/>); Spatial History Project, CESTA, Stanford University (<http://web.stanford.edu/group/spatialhistory/cgi-bin/site/index.php>); Roy Rosenzweig Center for History and New Media, George Mason University (<https://rrchnm.org/>); Scholars' Lab, University of Virginia (<http://scholarslab.org/>); Institute for Advanced Technology in the Humanities, University of Virginia (<http://www.iath.virginia.edu/>); Maryland Institute for Technology in the Humanities, University of Maryland (<https://mith.umd.edu/>); MIT HyperStudio, Massachusetts Institute of Technology (<http://hyperstudio.mit.edu/>); and Matrix, Michigan State University (<https://www2.matrix.msu.edu/>).

VI. Conclusion

The Working Group believes supporting digital scholarship is a natural and inevitable extension of the work the Library has done to acquire, describe, preserve, secure, and provide access to a universal collection of knowledge. A greater investment in collections readiness, community, and services is required not only to meet current demand but to position the Library to become an ever more vital resource in the 21st century.